



## **Metagenomic analyses reveal phylogenetic diversity of carboxypeptidase gene sequences in activated sludge of a wastewater treatment plant in Shanghai, China**

Jin, Hao; Li, Bailin; Peng, Xu; Chen, Lanming

*Published in:*  
Annals of Microbiology

*DOI:*  
[10.1007/s13213-013-0704-z](https://doi.org/10.1007/s13213-013-0704-z)

*Publication date:*  
2014

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Jin, H., Li, B., Peng, X., & Chen, L. (2014). Metagenomic analyses reveal phylogenetic diversity of carboxypeptidase gene sequences in activated sludge of a wastewater treatment plant in Shanghai, China. *Annals of Microbiology*, 64(2), 689-697. <https://doi.org/10.1007/s13213-013-0704-z>

# Metagenomic analyses reveal phylogenetic diversity of carboxypeptidase gene sequences in activated sludge of a wastewater treatment plant in Shanghai, China

Hao Jin · Bailin Li · Xu Peng · Lanming Chen

Received: 4 January 2013 / Accepted: 1 August 2013 / Published online: 28 August 2013  
© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** Activated sludge of wastewater treatment plants carries a diverse microflora. However, up to 80–90 % of microorganisms in activated sludge cannot be cultured by current laboratory techniques, leaving an enzyme reservoir largely unexplored. In this study, we investigated carboxypeptidase diversity in activated sludge of a wastewater treatment plant in Shanghai, China, by a culture-independent metagenomic approach. Three sets of consensus degenerate hybrid oligonucleotide primers (CODEHOPs) targeting conserved domains of public carboxypeptidases have been designed to amplify carboxypeptidase gene sequences in the metagenomic DNA of activated sludge by PCR. The desired amplicons were evaluated by carboxypeptidase sequence clone libraries and phylogenetic analyses. We uncovered a significant diversity of carboxypeptidases present in the activated sludge. Deduced carboxypeptidase amino acid sequences (127–208 amino acids) were classified into three distinct clusters,  $\alpha$ ,  $\beta$ , and  $\gamma$ . Sequences belonging to clusters  $\alpha$  and  $\beta$  shared 58–97 % identity to known carboxypeptidase sequences from diverse species, whereas sequences in the cluster  $\gamma$  were remarkably less related to public carboxypeptidase homologous in the GenBank database, strongly suggesting that novel carboxypeptidase families or microbial niches exist in the activated sludge. We also observed numerous carboxypeptidase sequences that were much closer to those from representative strains present in industrial and sewage

treatment and bioremediation. Thermostable and halotolerant carboxypeptidase sequences were also detected in clusters  $\alpha$  and  $\beta$ . Coexistence of various carboxypeptidases is evidence of a diverse microflora in the activated sludge, a feature suggesting a valuable gene resource to be further explored for biotechnology application.

**Keywords** Carboxypeptidase · Phylogenetic diversity · Activated sludge · Culture-independent approach

## Introduction

Activated sludge is used for treating sewage and industrial wastewaters, and it is the most common biological wastewater treatment technology used in industrialized countries (Kim et al. 2010). Sewage and industrial wastewater contain high amounts of complex organic matter which is mainly in the form of proteins and lipids (Raunkjær et al. 1994), and often also nitrogen (as ammonium and organic nitrogen) and phosphorus. The biomass that reduces the complex content of the sewage is the activated sludge, biological flocs mainly composed of saprotrophic bacteria, and some protozoa and small metazoans (Wilén et al. 2008). Microorganisms not only take up some small molecules for intracellular metabolism but also enzymatically hydrolyze a large fraction of the organic matter through a series of hydrolytic reactions to smaller units, which can be taken up by the bacterial cell uptake system (Gessesse et al. 2003). Previous research indicated that activated sludge is a good source for studying and discovering proteases (Gessesse et al. 2003), which play a vital role in the extracellular catabolism of organic matter in activated sludge. As one important group of proteases, carboxypeptidases (EC number 3.4.16–3.4.18) hydrolyze a peptide bond of amino acid residues at the carboxyl-terminal (C-terminal) end of a protein or peptide. They have diverse functions ranging from catabolism to protein post-translational modification and regulation of biological

H. Jin · B. Li · L. Chen (✉)

Key Laboratory of Quality and Safety Risk Assessment for Aquatic Products on Storage and Preservation (Shanghai), China Ministry of Agriculture, Engineering Centre for Quality Control and Risk Assessment of Aquatic Products, College of Food Science and Technology, Shanghai Ocean University, 999 Hu Cheng Huan Road, Shanghai 201306, People's Republic of China  
e-mail: lmchen@shou.edu.cn

X. Peng

Department of Biology, Copenhagen University, Ole Maaløes Vej 5, 2200 N Copenhagen, Denmark

processes (Neurath and Walsh 1976). Based on their active sites of catalytic reactions, carboxypeptidases are usually classified into metallo-carboxypeptidases (EC number 3.4.17), serine carboxypeptidases (EC number 3.4.16), and cysteine carboxypeptidase (or thiol carboxypeptidases) (EC number 3.4.18).

However, up to 80–90 % of the microorganisms detected in activated sludge by 16S rRNA genes based on molecular taxonomic studies cannot be cultured using standard cultivation techniques (Amann et al. 1995), leaving a potentially valuable resource largely unexplored. Recently, metagenomic technology has been used successfully to obtain microbial products from uncultivated microorganisms in various environments (Amann et al. 1995; Chu et al. 2008; Lee et al. 2008). Genes encoding novel enzymes, such as cellulases, chitinases, amylases, and lipases, have been screened from metagenomic libraries (Cottrell et al. 1999; Ferrer et al. 2005; Hong et al. 2007). Nevertheless, little if anything has been done to systematically investigate carboxypeptidases in microbial communities of complex environments. In this study, we focused on the diversity of carboxypeptidases in activated sludge of wastewater treatment systems in Shanghai, China, by a culture-independent metagenomic approach. Based on conserved domains including active site signature sequences of known carboxypeptidases in the GenBank database, a set of CODEHOPs was designed to amplify carboxypeptidase gene sequences in the metagenomic DNA of activated sludge by polymerase chain reaction (PCR). Our results provide the first evidence of a substantial level of carboxypeptidase sequence diversity in the microbial community of activated sludge in industrial and municipal wastewater treatment.

## Materials and methods

### Bacterial strains, plasmids and culture conditions

*Escherichia coli* TOP10 [genotype: F<sup>−</sup> *mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ80 *lacZ* ΔM15 Δ*lacX*74 *recA1* *araD*139 Δ(*ara-leu*)7697 *galU* *galK* *rpsL* (St<sup>R</sup>) *endA1-nupG*] (TianGen Biotech, Beijing, China) was used as a host strain for transformation of recombinant plasmids in the cloning. Plasmid pGM-T (TianGen) was employed as a cloning vector for PCR products. *E. coli* was routinely grown in Luria-Bertani (LB) medium (Sambrook and Russell 2001) aerobically at 37 °C, while recombinant *E. coli* strains were grown in LB medium in both liquid broth and agar plates supplemented with ampicillin (100 μg/ml).

### Extraction of metagenomic DNA from activated sludge

Samples of activated sludge were collected from a wastewater treatment plant located in Lingang, Pudong District, Shanghai, China, where industrial and municipal wastewater was treated.

Extraction of metagenomic DNA was carried out immediately after samples had been transported on ice to the laboratory in Shanghai Ocean University (Shanghai, China). Metagenomic DNA was isolated according to the method of in situ lysis of microorganisms described by Zhou et al. (1996) with some modifications. Briefly, a sample (50 g) of wet activated sludge and 135 ml DNA extraction buffer [100 mmol/l Tris-HCl, 100 mmol/l EDTA, 1.5 mol/l NaCl and 1 % (w/v) hexadecyltrimethylammonium bromide, pH 8.0; Sigma-Aldrich, MO, USA] were blended vigorously at 37 °C for 0.5 h. Then, 0.5 ml protease K (20 mg/ml) and 2.0 ml SDS (20 %) were added, and the mixture was incubated at 50 °C for 3 h with gentle end-over-end inversions every 20 min. The supernatant was collected after centrifugation at 6,000 g for 10 min at 25 °C in a high speed centrifuge (Hitachi Himac CR 21G; Hitachi Koki, Japan), and then mixed with an equal volume of chloroform:isopropyl alcohol (24:1). The aqueous phase was recovered by centrifugation at 12,000 g at 4 °C for 10 min, in which DNA was precipitated by adding 0.1 volume of 3 M sodium acetate (pH 5.3) and 2 volumes of chilled anhydrous ethanol, held at −20 °C for 2 h, and then recovered by centrifugation at 13,000 g for 15 min at 4 °C. The DNA pellet was washed with 70 % ethanol, and resuspended with 200 μl sterile Milli-Q water (Millipore, Billerica, MA, USA). DNA concentration was determined in a SAM 1000 Spectrophotometer (Merinton Technology, Beijing, China). The isolated DNA was further purified by using the Wizard DNA Clean-Up System (Promega, WI, USA), according to the manufacturer's instruction.

### Designing of degenerate primers for PCR

Degenerate primers were designed based on public carboxypeptidase amino acid sequences in the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/Entrez>). By using the program Clustal W2 (Thompson et al. 1994) (<http://www.ebi.ac.uk/Tools/msa/ClustalW2/>), multiple sequence alignments were carried out with each of the three sets of carboxypeptidase sequences which were 490–511, 491–506, and 391–404 amino acids, respectively, representing different taxonomic lineages. The resulting conserved blocks of amino acid residues were used as input to load into the program iCODEHOP (Boyce et al. 2009) (<https://icodehop.cphi.washington.edu/i-codehop-context/Welcome>), by which degenerate primers were designed. A CODEHOP is a hybrid primer consisting of a 3'-degenerate 'core' and 5'-non degenerate 'clamp' region (Boyce et al. 2009). The default values of iCODEHOP were taken. The maximum degeneracy of 3'-'core' region of each CODEHOP was set at 128, while the annealing temperature of 5'-non degenerate 'clamp' of CODEHOP was set at 60 °C. Three sets of degenerate primers were chosen for PCR reactions in this study, the sequences of which were as following:

CG-F: 5'-GTCCGTGCACCCCtttygyggngg-3' CG-R: 5'-CCAGGGTGTAGCAGGgraartancra-3', CFC-F: 5'-GGTCCTGAAGCAGATCggtaygaytt-3', CFC-R: 5'-GCCCAGGGCGTAGswnggraarta-3', CFA-F: 5'-CCGGGCCGACATCgaygcnyncc-3', and CFA-R: 5'-GGTGGGGCATGGAGscrtgncnc-3'. Oligonucleotide primers were synthesized by the Shanghai Sangon Biological Engineering Technology Services (Shanghai, China).

#### PCR conditions

The carboxypeptidase sequences were amplified by PCR using the purified metagenomic DNA of the activated sludge as templates. PCR amplification was performed in a 20- $\mu$ l reaction volume containing 1 $\times$  Premix Ex Taq Version 2.0 (Japan TaKaRa BIO, Dalian, China), 2.5  $\mu$ M each of the oligonucleotide primers, and approximately 10–20 ng of template DNA. All amplifications were performed in a Peter Thermal Cycler (BIO-RAD, CA, USA).

Touchdown PCR was performed with the degenerate primer pair CG-F and CG-R under the following conditions: initial denaturation of 94 °C for 5 min was followed by 15 cycles consisting of denaturation at 94 °C for 30 s, primer annealing at 65 °C for 30 s with a gradual decrease of 1 °C per cycle, and elongation at 72 °C for 1 min, and then followed by 20 cycles consisting of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min. The final extension step was at 72 °C for 10 min. The best annealing temperature of the degenerate primer pair CFC-F and CFC-R was determined by temperature gradient PCR, and established where a single band was observed on a 1.2 % agarose gel under the following conditions: initial denaturation of 95 °C for 5 min was followed by 30 cycles consisting of denaturation at 94 °C for 30 s, primer annealing at 47.2 °C for 30 s, and elongation at 72 °C for 1 min, followed by final elongation at 72 °C for 10 min. Similarly, the temperature gradient PCR was also established with the degenerate primer pair CFA-F and CFA-R under the similar conditions as described above, except the primer annealing at 58.3 °C for 30 s, and elongation at 72 °C for 45 s. The quality of the extracted DNA was tested by PCR with bacterial 16S rDNA universal primer pair 27 F (5'-AGAGTTTGTATCCTGGCTCAG-3') and 1492R (5'-TACCTTGTACGACTT-3'). PCR reactions were carried out as following: initial denaturation of 95 °C for 5 min was followed by 30 cycles consisting of denaturation at 94 °C for 30 s, primer annealing at 55 °C for 30 s, and elongation at 72 °C for 2 min, and a final extension of 72 °C for 10 min.

A sample (5  $\mu$ l) of each amplification reaction was analyzed by agarose gel electrophoresis running in a Sub-Cell GT cell (BIO-RAD) with a 1.2 % or 0.7 % agarose gel and 1 $\times$  Tris-acetate-EDTA (TAE) buffer (40 mM Tris-acetic acid and 2 mM ethylenediaminetetraacetate disodium, EDTA- $\text{Na}_2$ ). After the run, agarose gels were stained with ethidium bromide (0.33 mg/l) for 20 min. Amplified DNA fragments were

visualized under short-wavelength UV light (260 nm) and imaged by the UVP EC3 Imaging System (UVP, CA, USA).

#### Construction of carboxypeptidase sequence clone libraries

After PCR amplification, the desired products with 5'-A overhangs yielded by the Taq DNA polymerase were purified by agarose gel electrophoresis, and recovered from the gel by Axygen Gel Extraction Kit (Axygen, CA, USA). The purified PCR products were individually measured for DNA concentration, and ligated into the pGM-T vector with 3' T overhangs at the insert site. The ligation reaction was preformed in a 10- $\mu$ l reaction volume with a molar ratio of vector: insert of 1:3, and incubated at 16 °C overnight, according to the manufacturer's instruction. Competent *E. coli* TOP 10 cells were transformed with the ligation DNA by the heat-shock method (Sambrook and Russell 2001). Briefly, 2  $\mu$ l of each ligation DNA was added into 100  $\mu$ l *E. coli* TOP 10 competent cells in a sterile 1.5-ml microcentrifuge tube, and incubated on ice for 30 min. The cells were then heat-shocked in water bath at 42 °C for 45 s, and immediately returned the tube to ice for 2 min. Subsequently, 900  $\mu$ l prewarmed LB medium was added to suspend the cells, followed by incubation at 37 °C for 1 h with shaking. A sample of 100  $\mu$ l of each transformation culture was spread onto LB-ampicillin plates containing 40  $\mu$ l 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-gal, 20 mg/ml) and 16  $\mu$ l isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG, 50 mg/ml). Plates were incubated overnight at 37 °C, and white colonies were randomly picked for further analysis. Transformation control was prepared with 0.1 ng of uncut plasmid pUC19.

#### Sequencing and phylogenetic analyses

The white recombinant colonies were randomly picked for plasmid DNA preparation by using the MiniBEST plasmid purification kit (TaKaRa, China). Automated DNA sequencing of the insertions on purified recombinant vectors was carried out using the ABI 3730XL capillary sequencer (Applied Biosystems, CA, USA) and BigDye<sup>®</sup> terminator version 3.1 cycle sequencing kit (Perkin-Elmer, MA, USA) at the China Human Genome Centre (Shanghai, China). The universal sequencing primers, forward (5'-TGTA AACGACGGCCAGT-3') and reverse (5'-CAGGAA ACAGCTATGACC-3'), on the pGM-T vector were utilized in the sequencing reactions. The average length of high quality sequences is 650–700 bp.

Raw DNA sequence data from both DNA strands were analyzed by using the program BioEdit (v.7.0.9, <http://www.mbio.ncsu.edu/BioEdit>). The sequencing reads were processed to trim usual poor-quality sequence at the 3' end of each read. The reads from both DNA strands were overlapped and assembled into full length contigs.

Individual consensus sequence was exported into a FASTA format, and aligned using the Clustal W2. Putative functions were inferred by using the Basic Local Alignment Search Tool (BLAST) (<http://www.ncbi.nlm.nih.gov/BLAST>). The neighbor-joining method in the molecular evolutionary genetics analysis software MEGA4 (v.4.0) (Tamura et al. 2007) was used to construct a phylogenetic tree. A bootstrap analysis with 1,000 replicates was carried out to check the robustness of the tree.

#### Nucleotide sequence accession numbers

The representative carboxypeptidase sequences obtained in this study have been deposited in the GenBank sequence database under the accession numbers from JN811663 to JN811672.

## Results and discussion

#### Degenerate primers targeting conserved domains of carboxypeptidases

The carboxypeptidase sequences were searched in the NCBI database, three sets of which were used for designing the CODEHOPs. A multiple alignment of eight to ten public carboxypeptidase sequences was carried out for each set by using the Clustal W2. Two blocks of highly conserved amino acid residues HPFCG and FGYFP, GYDF and GYFP, and DALPI and GGH were identified within the multiple alignments, respectively (Fig. 1). A set of CODEHOPs was yielded after individual Clustal W2 alignment was input into the iCODEHOP. The relative quality of individual CODEHOPs was further analyzed based on the meta-data about each primer pool created by the iCODEHOP. The degenerate primer pairs CG-F and CG-R, CFC-F and CFC-R, and CFA-F and CFA-R were chosen to amplify carboxypeptidase sequences from the metagenomic DNA of the activated sludge, with an expected size of 582, 626, and 396 bp, respectively.

#### Cloning of carboxypeptidase gene sequences from the metagenomic DNA extracts

The quality of the extracted metagenomic DNA was tested by PCR with the bacterial 16S rDNA universal primers 27F and 1492R. A single clear band of approximately 1.5 kb was detected by electrophoresis on a 0.7 % agarose gel, indicating good quality of the isolated metagenomic DNA from activated sludge (data not shown).

The touchdown PCR reactions were performed with the degenerate primer pair CG-F and CG-R, and the desired amplicon with a length of near 0.6 kb was observed by electrophoresis on a 1.2 % agarose gel (Fig. 2). An amplicon

in size between 0.6 and 0.7 kb was obtained by temperature gradient PCR with the degenerate primer pair CFC-F and CFC-R at an annealing temperature of 47.2 °C, similarly a 0.4-kb product was also obtained with the CFA-F and CFA-R at 58.3 °C (Fig. 2).

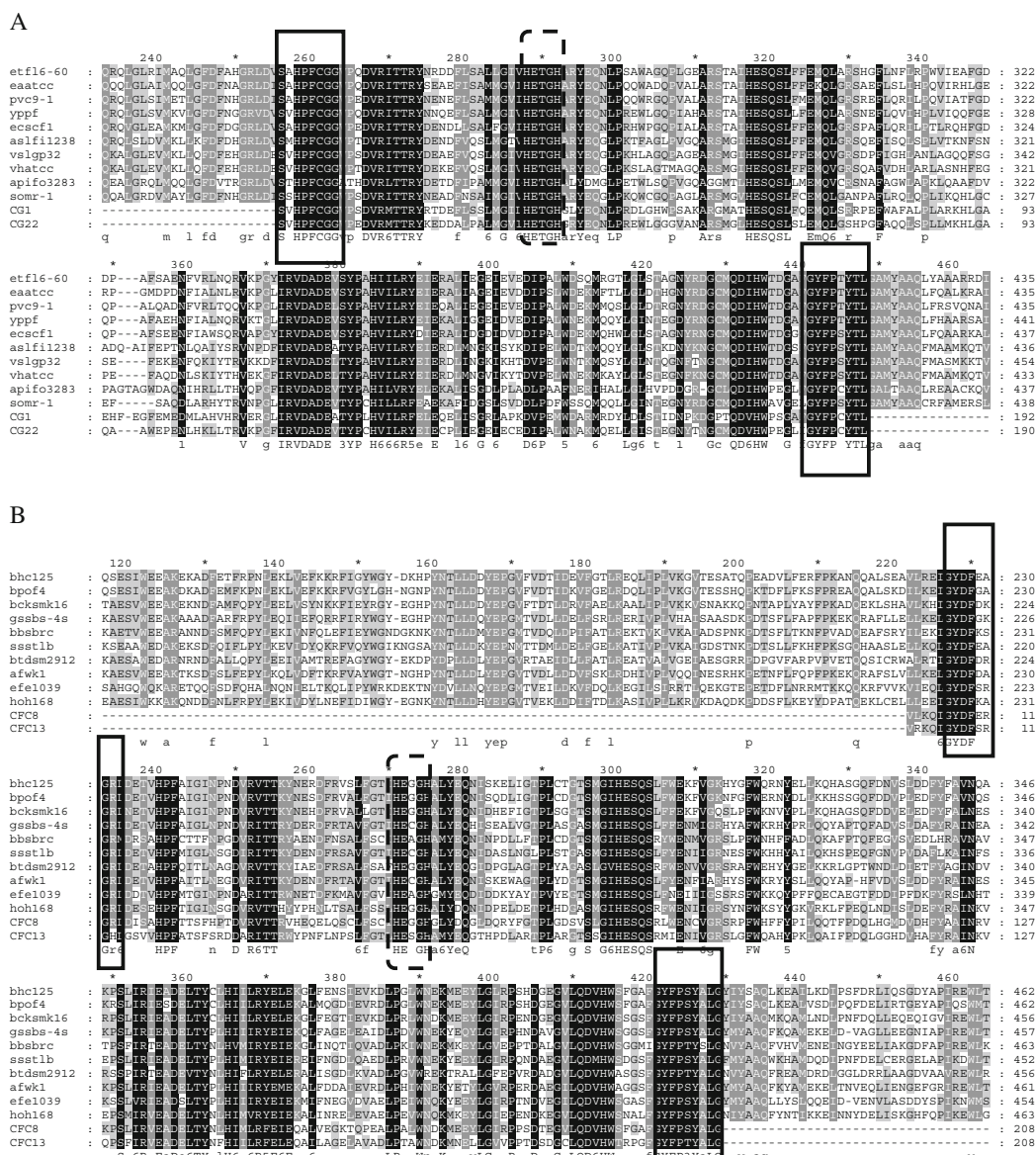
Each of three amplicons was purified, ligated into the pGM-T vector, and transformed into *E. coli* TOP10 competent cells, which yielded three carboxypeptidase sequence clone libraries, designated as CG, CFC, and CFA, respectively. About 50 white colonies from each of the three libraries were randomly picked for plasmid preparation and sequencing analyses. More than 90 % of recombinant vectors of the three libraries carried inserts in size similar to that of their corresponding amplicons by agarose gel electrophoresis analyses (data not shown), indicating the validity of the constructed libraries.

#### Diversity of carboxypeptidase gene sequences in the microbial community of the activated sludge

In total, 102 sequence contigs in high quality were obtained from the three libraries. The evolutionary relationships between these sequences and public carboxypeptidase were assessed by phylogenetic analyses. Based on the deduced amino acid sequences ranging from 127 to 208 residues, a phylogenetic tree was constructed by using the MEGA4, and it revealed three distinct clusters, designated as clusters  $\alpha$ ,  $\beta$ , and  $\gamma$  (Fig. 3). Cluster  $\alpha$  contained sequences derived from the libraries CG and CFC, whereas cluster  $\beta$  consisted of sequences solely from the library CFA, and cluster  $\gamma$  was composed of sequences from all three libraries.

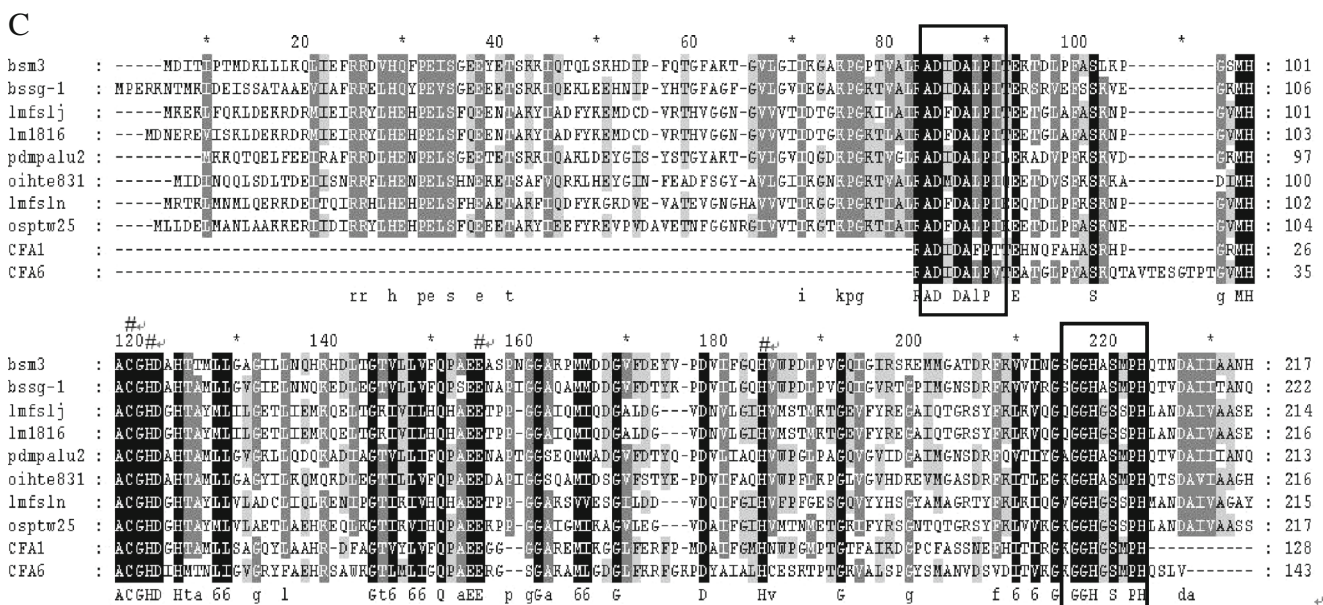
More than half of the sequences fell into cluster  $\alpha$ , the majority of which were yielded from the library CG, while about one-fourth were from the library CFC. All the public carboxypeptidase sequences used for designing the degenerate primers CG-F/R and CFC-F/R were distributed in cluster  $\alpha$ , but belonged to two notably different subclusters (Fig. 3). In cluster  $\alpha$ , sequences from the library CG shared 68–97 % identity at the amino acid level with carboxypeptidase sequences from different genera including *Nitrosomonas*, *Leptothrix*, *Aeromonas*, and *Mesorhizobium*, while the sequences from the library CFC showed 57–84 % identity to the homologous from *Candidatus Nitrospira*, *Chitinophaga*, *Herpetosiphon*, *Anaerolinea*, *Rhodothermus*, and *Planctomyces*. There was a significant number of amino acid residues conserved in the multiple sequence alignments with sequences in cluster  $\alpha$ , which spanned a common active site sequence motif His-Glu-Xaa-Xaa-His (HEXXH) (Fig. 1a, b), characteristic of the peptidase M32 family, a subclass of metallo-carboxypeptidases distributing mainly in bacteria and archaea. The two histidines coordinated a divalent cation such as  $Zn^{2+}$  or  $Co^{2+}$ , while a downstream glutamate polarized and oriented a zinc-coordinating water molecule during catalysis (Lee et al. 2009; Matthews et al. 1974). In addition, in cluster  $\alpha$ , the sequences





**Fig. 1** Alignments of carboxypeptidase amino acid sequences from homologues in the GenBank, and representative novel sequences obtained in this study by PCR amplification of metagenomic DNA from activated sludge samples with newly designed degenerate primers. The numbering of sequences is indicated above the alignments. The starting and ending residue numbers of each protein sequence used in this study are indicated for each sequence. Conserved amino acid residues are highlighted and shown below each alignment, and those for designing CODEHOPs are boxed with solid lines. Common active site sequence motifs are boxed with dotted lines, and conserved feature residues are located with hash marks above the aligned sequences. The carboxypeptidase sequences displayed in the alignments are as following: **a** *Edwardsiella tarda* FL6-60 (ADM41696.1; etf16-60), *Erwinia amylovora* ATCC 49946 (YP\_003538882.1; eaatcc), *Pantoea vagans* C9-1 (YP\_003931068.1; pvc9-1), *Yersinia pestis* Pestoides F (ABP39239.1; yppf), *Enterobacter cloacae* SCF1 (ADO48513.1; ecscf1), *Aliivibrio salmonicida* LFI1238 (YP\_002263204.1; aslf1238), *Vibrio splendidus* LGP32 (YP\_002416903.1; vs1gp32), *Vibrio parvulus* ATCC BAA-1116 (YP\_001445666.1; vhatcc), *Acetobacter pasteurianus*

IFO 3283-01 (YP\_003188184.1; apifo3283), *Shewanella oneidensis* MR-1 (NP\_716995.1; somr-1). **b** *Bacillus halodurans* C-125 (NP\_243015.1; bhc125), *Bacillus pseudofirmus* OF4 (YP\_003428150.1; bpof4), *Bacillus clausii* KSM-K16 (BAD64530.1; bcksmk16), *Geobacillus* sp. SBS-4S (BAH28805.1; gsbs-4 s), *Brevibacillus brevis* NBRC 100599 (YP\_002771690.1; bbsbrc), *Solibacillus silvestris* StLB046 (BAK18141.1; ssstlb), *Bacillus tusciae* DSM 2912 (YP\_003589815.1; btdsm2912), *Anoxybacillus flavithermus* WK1 (ACJ33726.1; afwk1), *Enterococcus faecium* E1039 (ZP\_06674096.1; efe1039), *Halothermothrix orenii* H 168 (ACL69173.1; hoh168). **c** *Bacillus* sp. m3-13 (ZP\_07707682.1; bsm3), *Bacillus* sp. SG-1 (ZP\_01858539.1; bssg-1), *Listeria monocytogenes* FSL J1-175 (ZP\_05388672.1; lmfs1j), *Listeria monocytogenes* 1816 (EGF39295.1; lm1816), *Planococcus donghaensis* MPA1U2 (ZP\_08095878.1; pdmpa1u2), *Oceanobacillus iheyensis* HTE831 (NP\_693505.1; oihte831), *Listeria monocytogenes* FSL N1-017 (ZP\_07073423.1; lmfs1n), *Ornithinibacillus* sp. TW25 (ZP\_08783624.1; osptw25). The CG1 and CG22, CFC8 and CFC13, CFA1 and CFA6 were obtained from CG, CFC, and CFA libraries in this study, respectively

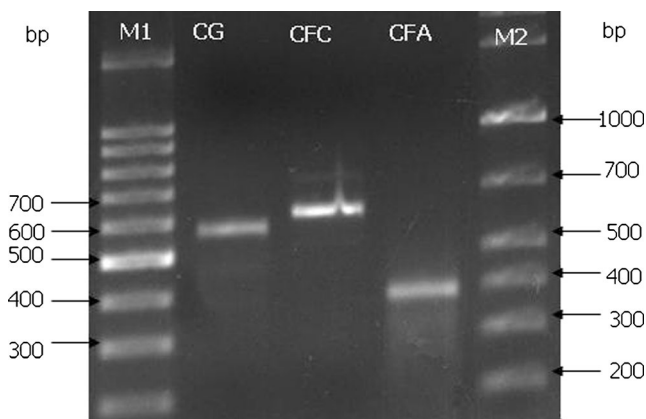


**Fig. 1** (continued)

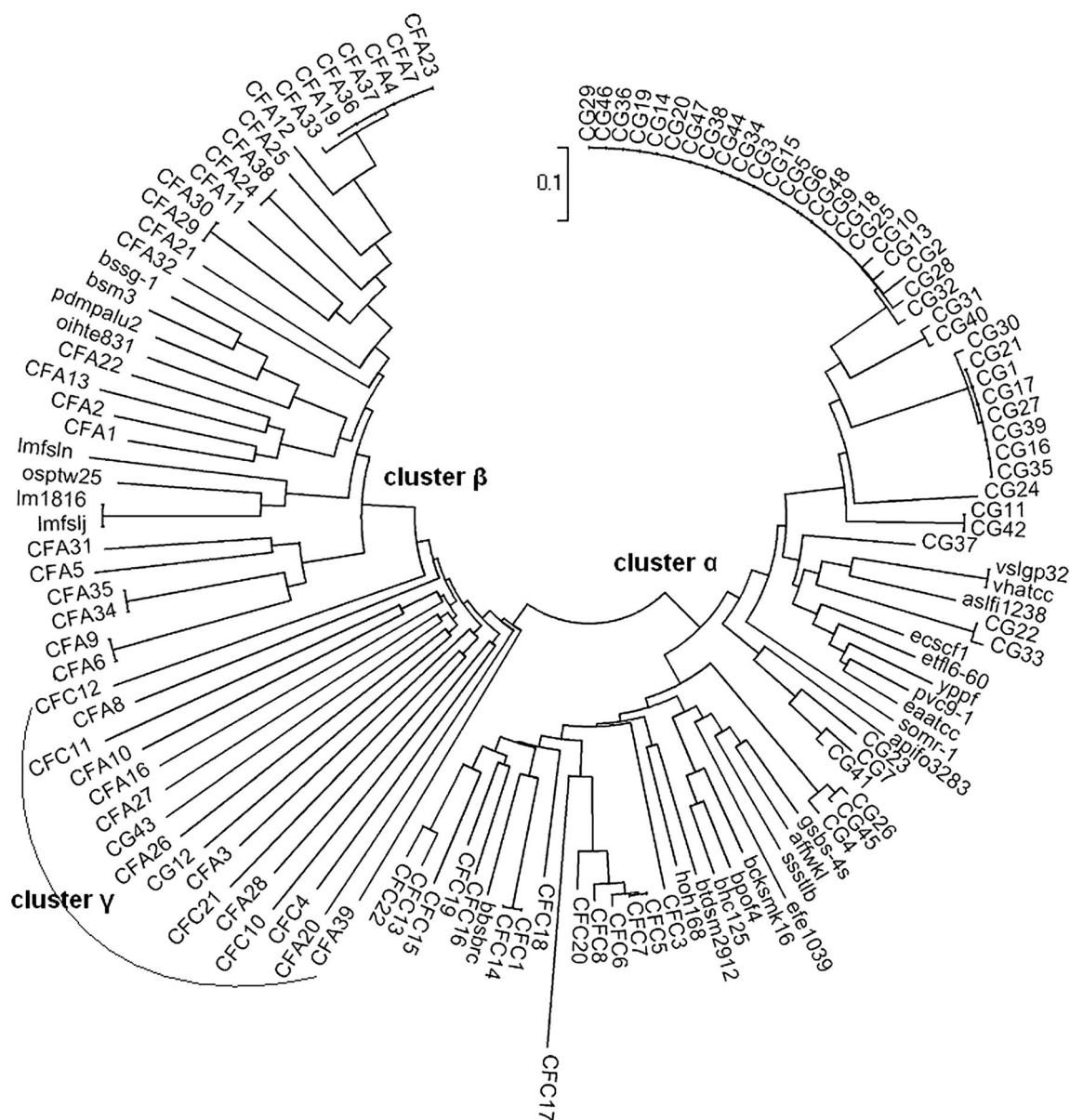
CFC1, CFC5, CFC6, CFC7, CFC8 (JN811665), CFC14, CFC17, and CFC20 displayed 59–84 % identity to the carboxypeptidase of a nitrite-oxidizing bacterium *Candidatus Nitrospira defluvii*, which was a key nitrifier in biological wastewater treatment of industrially contaminated sites and wastewater (Kostan et al. 2010). The CG4, CG26 (JN811671), and CG45 were closely related to the carboxypeptidase of an ammonia-oxidizing *Nitrosomonas* sp. with identities of 92 %, while the CFC13 (JN811666) and CFC22 were closely related to that from a nature's scavenger, *Herpetosiphon aurantiacus* (Quinn and Skerman 1980), with similarity of 60 and 62 %, respectively. The CG22 (JN811664) and CG33 shared 68 and 73 % identity to the

carboxypeptidase of an aerobic sheath-forming bacterium *Leptothrix cholodnii* that was often found in oligotrophic and metal-rich aquatic environments (Takeda et al. 2005). These results underline the considerable phylogenetic diversity of the carboxypeptidase gene sequences in the microbial community of the activated sludge.

Cluster  $\beta$  was grouped by one-fourth of the sequences, all of which were derived from the library CFA, and shared 58–78 % identity to proteases from different species, except for two sequences with less than 50 % coverage. In cluster  $\beta$ , most sequences were related to the peptidase M20D family, amidohydrolase, and one to the M21D family of the genera including *Haliscomenobacter*, *Runella*, *Mucilaginibacter*, *Chitinophaga*, *Comamonas*, *Cyclobacterium*, *Acidovorax*, *Bordetella*, *Chthoniobacter*, *Ruegeria*, and *Dyadobacter*. In addition, the CFA9 and CFA5 were close to the carboxypeptidase Ss0 subfamily protein of *Myxococcus* (YP\_004669645.1), the CFA21 to the bifunctional carboxypeptidase of *Thermococcus* (YP\_002306730.1), and the CFA32 to the thermostable carboxypeptidase1 of *Thermomicrobium* (YP\_002523939.1). Similar to cluster  $\alpha$ , the public carboxypeptidase sequences used for designing the CFA-F/R primers were spread in the cluster  $\beta$ , and numerous conserved residues were also observed in the multiple sequence alignment containing the metal binding sites of the M20 family conserved domains (Fig. 1c). The peptidase M20 family including carboxypeptidases belongs to zinc peptidases which play vital roles in metabolic and signaling pathways throughout all kingdoms of life (Rawlings and Barrett 1995). Moreover, in cluster  $\beta$ , the sequences CFA33, CFA34, CFA13, and CFA1 (JN811667) showed 63–75 % identities to the peptidases of *Haliscomenobacter*, *Comamonas*, and *Acidovorax* species



**Fig. 2** Agarose gel electrophoresis analysis of PCR products amplified by degenerate primers with metagenomic DNA isolated from activated sludge samples as templates. *M1* 1-kb plus DNA ladder, *M2* 100-bp DNA ladder. The CG, CFC, and CFA are amplicons yielded by the degenerate primers CG-F and CG-R, CFC-F and CFC-R, and CFA-F and CFA-R, respectively



**Fig. 3** Phylogenetic tree of deduced carboxypeptidase amino acid sequences obtained from microbial community of activated sludge samples in industrial and municipal wastewater. Based on the 102 novel sequences ranged in size from 127 to 208 amino acid residues, the neighbor-joining

tree was constructed by the MEGA 4.0 together with a selected set of the public carboxypeptidase sequences. The length of the branches indicates the divergence among the amino acid sequences

that have been commonly observed in activated sludge samples originating from both municipal and industrial wastewater treatment plants (Kragelund et al. 2008; Kraigher et al. 2008; Srinandan et al. 2011). These results strongly reinforce the high diversity of carboxypeptidases present in the activated sludge.

A total of 16 % of the sequences fell into cluster γ, about half of which were generated from the library CFA, while the others were from the libraries CFC and CG. In contrast to the high identity and coverage (72–99 %) in clusters α and β, the deduced amino acid sequences in cluster γ remarkably

displayed low levels of similarity and coverage to public homologues and hypothetical proteins in GenBank, except for CFA20, CFA8, and CFC4. CFA20 carried 48 % identity to the ATP-dependent protease La (Lon) domain protein of 218 amino acids from *Stigmatella aurantiace* DW4/3-1 (ZP\_01466891.1), and CFA8 (JN811669) had 64 % identity to the glycoprotease family metalloendopeptidase of *Thermobaculum terrenum* (YP\_003321788.1), while CFC4 (JN811670) showed 88 % identity to the hypothetical protein NIDE4078 of *Candidatus Nitrospira defluvii* (Lücker et al. 2010). Interesting, in the cluster γ, some sequences derived



from the three libraries notably had 7–30 more amino acids than the desired products amplified with three CODEHOPs, while the other sequences were slightly shorter than predicted lengths. No public carboxypeptidase sequences were grouped into cluster  $\gamma$ , but a common ancestor was shared among the three clusters (Fig. 3). Moreover, sequences in cluster  $\gamma$  seemed less close to each other than to database homologues, strongly suggesting that novel carboxypeptidase families or microbial niches existed in the activated sludge.

**Specificity of the CODEHOPs-based PCR for screening novel carboxypeptidases from complex environmental samples**

The diversity and presence of different conserved motifs among carboxypeptidase families led to the fact that no primers are omnipotent for broad range-based amplification of carboxypeptidase gene sequences. This problem may be solved by designing more sets of primers specific for carboxypeptidase families. Nevertheless, the metagenomic technique is a powerful tool for evaluating phylogenetic diversity of carboxypeptidases present in activated sludge by a culture-independent approach. In this study, a large number of carboxypeptidase sequences were cloned by PCR amplification of metagenomic DNA of the activated sludge samples. The inferred amino acid sequences in three libraries had various identities to public carboxypeptidase sequences of multifarious bacterial resources, including representative strains present in the process of industrial and sewage treatment and bioremediation. In addition, the CG1 (JN811663) and CFC19 in cluster  $\alpha$  shared 72 and 60 % identity to the thermostable carboxypeptidases of *Mesorhizobium loti* (NP\_104928.1) and *Rhodothermus marinus* (YP\_003290669.1), respectively, while the CFA6 (JN811668) and CFA9 in cluster  $\beta$  was related to that of the halotolerant *Myxococcus fulvus* (YP\_004669645.1). Interesting, the CFA2 in the cluster  $\beta$  and the CG37 (JN811672) in cluster  $\alpha$  were closely related to the carboxypeptidases of the pathogenic bacteria belonging to *Bordetella* and *Aeromonas* species with identities of 72 and 97 %, respectively; the latter was also revealed in the activated sludge samples by 454 pyrosequencing (Ye and Zhang 2011). Moreover, some sequences in clusters  $\alpha$  and  $\beta$  had 63–76 % identity to the carboxypeptidases of *Mucilaginibacter* and *Chitinophaga* species that are able to degrade pectin, xylan, chitin, laminarin, and some other polysaccharides. Taken together, our data revealed a significant diversity of carboxypeptidase sequences existing in the metagenomic DNA of the activated sludge samples, suggesting a diverse microflora in this special environment.

To our knowledge, this study constitutes the first investigation of microbial carboxypeptidase diversity in complex environments by using a culture-independent strategy. The multifunctional property of proteases is based on their diverse

structures, which is a consequence of differences in protein sequences. Thus, screening for novel carboxypeptidases with various characteristics is of interest from both industrial and academic standpoints. The results in this study demonstrated the specificity and effectiveness of the sequence-based screening approach for identifying novel carboxypeptidases directly from complex environmental samples, but also expanded the views on the biodiversity and utilization of activated sludge in biotechnology application.

**Acknowledgments** This work was supported by Grant No.09320503600 from Shanghai Municipal Science and Technology Commission and Grant No.B-9500-10-0004 from Shanghai Municipal Education Commission. We thank T. Zhang for the help in designing degenerate primers, and Y. Shi in extracting metagenomic DNA from the activated sludge samples.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Amann RI, Ludwig W, Schleifer K-H (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Res* 59:143–169
- Boyce R, Chilana P, Rose TM (2009) iCODEHOP: a new interactive program for designing Consensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. *Nucleic Acids Res* 37:222–228
- Chu X, He H, Guo C, Sun B (2008) Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl Microbiol Biotechnol* 80:615–625
- Cottrell MT, Moore JA, Kirchman DL (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* 65:2553–2557
- Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VA, Strompl C, Elborough K, Jarvis G, Neef A, Yakimov MM, Timmis KN, Golyshin PN (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* 7:1996–2010
- Gessesse A, Dueholm T, Petersen SB, Nielsen PH (2003) Lipase and protease extraction from activated sludge. *Water Res* 37:3652–3657
- Hong KS, Lim HK, Chung EJ, Park EJ, Lee MH, Kim JC, Choi GJ, Cho KY, Lee SW (2007) Selection and characterization of forest soil metagenome genes encoding lipolytic enzymes. *J Microbiol Biotechnol* 17:1655–1660
- Kim HS, Gellner JW, Boltz JP, Freudenberg RG, Gunsch CK, Schuler AJ (2010) Effects of integrated fixed film activated sludge media on activated sludge settling in biological nutrient removal systems. *Water Res* 44:1553–1561
- Kostan J, Sjöblom B, Maixner F, Mlynek G, Furtmüller PG, Obinger C, Wagner M, Daims H, Djinić-Carugo K (2010) Structural and functional characterisation of the chlorite dismutase from the nitrite-oxidizing bacterium “*Candidatus Nitrospira defluvi*”: identification of a catalytically important amino acid residue. *J Struct Biol* 172:331–342
- Kragelund C, Levantesi C, Borger A, Thelen K, Eikelboom D, Tandoi V, Kong Y, Krooneman J, Larsen P, Thomsen TR, Nielsen PH (2008) Identity, abundance and ecophysiology of filamentous bacteria belonging to the *Bacteroidetes* present in activated sludge plants. *Microbiology* 154:886–894

- Kraigher B, Kosjek T, Heath E, Kompare B, Mandic-Mulec I (2008) Influence of pharmaceutical residues on the structure of activated sludge bacterial communities in wastewater treatment bioreactors. *Water Res* 42:4578–4588
- Lee CM, Yeo YS, Lee JH, Kim SJ, Kim JB, Han NS, Koo BS, Yoon SH (2008) Identification of a novel 4 hydroxyphenylpyruvate dioxygenase from the soil metagenome. *Biochem Biophys Res Commun* 370:322–326
- Lee MM, Isaza CE, White JD, Chen RP, Liang GF, He HT, Chan SI, Chan MK (2009) Insight into the substrate length restriction of M32 carboxypeptidases: characterization of two distinct subfamilies. *Proteins* 77:647–657
- Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, Rattei T, Damsté JS, Spieck E, Le Paslier D, Daims H (2010) A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci USA* 107:13479–13484
- Matthews BW, Weaver LH, Kester WR (1974) The conformation of thermolysin. *J Biol Chem* 249:8030–8044
- Neurath H, Walsh KA (1976) Role of proteolytic enzymes in biological regulation. *Proc Natl Acad Sci USA* 73:3825–3832
- Quinn GR, Skerman VBD (1980) *Herpetosiphon*—nature’s scavenger? *Curr Microbiol* 4:57–62
- Raunkjær K, Hvitved-Jacobsen T, Nielsen PH (1994) Measurement of pools of protein, carbohydrate and lipid in domestic wastewater. *Water Res* 28:251–262
- Rawlings ND, Barrett AJ (1995) Evolutionary families of metallo-peptidases. *Methods Enzymol* 248:183–228
- Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*, 3rd edn. Cold Spring Harbor, New York
- Srinandan CS, Shah M, Patel B, Nerurkar AS (2011) Assessment of denitrifying bacterial composition in activated sludge. *Bioresour Technol* 102:9481–9489
- Takeda M, Makita H, Ohno K, Nakahara Y, Koizumi J (2005) Structural analysis of the sheath of a sheathed bacterium, *Leptothrix cholodnii*. *Int J Biol Macromol* 37:92–98
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Wilén B-M, Onuki M, Hermansson M, Lumley D, Mino T (2008) Microbial community structure in activated sludge floc analysed by fluorescence in situ hybridization and its relation to floc stability. *Water Res* 42:2300–2308
- Ye L, Zhang T (2011) Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. *Environ Sci Technol* 45:7173–7179
- Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62:316–322